To appear in: Bunt, H & R.J. Beun, Selected Papers from the Proceedings of the Second International Conference on Cooperative Multimodal Communication, Lecture Notes in Artificial Intelligence, Berlin: Springer-Verlag.

Relating Imperatives to Action

Paul Piwek

ITRI – Information Technology Research Institute, Brighton
Email: Paul.Piwek@itri.bton.ac.uk

November 20, 2000

Abstract. The aim of this paper is to provide an analysis of the use of logically complex imperatives, in particular, imperatives of the form 'Do A_1 or A_2' and 'Do A, if B'. I argue for an analysis of imperatives in terms of Classical Logic which takes into account the influence of background information on imperatives. I show that by doing so one can avoid some counter-intuitive results which have been associated with analyses of imperatives in terms of Classical Logic. In particular, I address Hamblin's observations concerning rule-like imperatives and Ross' Paradox. The analysis is carried out within an Agent-based Logical framework. This analysis explicates what it means for an agent to have a successful policy for action with respect to satisfying his or her commitments, where some of these commitments have been introduced as a result of imperative language use.

1 Introduction

Imperatives[1] are typically used to get an addressee to do something. As a result, a key notion in the analysis of imperatives is the notion of the *satisfaction of an imperative* (by its addressee). Understanding the meaning of an imperative involves knowing under which circumstances it has been satisfied. Various authors, notably Stenius (1967) and Searle (1969), have proposed that we analyse this

[1] I will use the word 'imperative' as a short-hand for 'the utterance of an imperative clause'. Imperative clauses are traditionally characterised by the lack of a subject, the use of a base form of the verb, and the absence of modals as well as tense and aspect markers. Examples of naturally occurring imperatives are: 'Get off the table', 'Don't forget about the deposit' and 'Hold on, are we late?'. This definition and the examples are taken from Biber et al. (1999:219).

notion of *satisfaction* by assuming that an imperative conveys both a proposition and the instruction to make this proposition true. For instance, the imperative 'Get off the table' conveys the proposition *that the addressee gets off the table* and the instruction that *the addressee makes it true that s/he gets off the table.* Whether or not an imperative has been satisfied can now be understood in terms of whether its propositional content has been made true. Although this account of imperatives sounds plausible, it has been shown to give rise to some apparently counter-intuitive results once it is formalised by means of Classical (Predicate) Logic. In this paper, I discuss a number of these results and demonstrate how they can be avoided by embedding the aforementioned type of analysis in an Agent-based model which takes into account the influence of background information on the use of imperatives.

2 The Puzzle of Logically Complex Imperatives

Let me indicate how a formalisation of the aforementioned analysis of imperatives appears to give rise to counter-intuitive results. In such a formalisation, the propositional content of imperatives is represented by means of Predicate Logical sentences. More precisely, there is a function f_C which maps an imperative clause, given some utterance context C, to a Predicate Logical representation of its propositional content. In particular, the natural language connective 'or' is rendered by means of the logical disjunction '$\vee$' and the natural language construction '. . ., if . . .' is represented by means of the material conditional '$\rightarrow$':

(1) $f_C(A_1 \textit{ or } A_2) = f_C(A_1) \vee f_C(A_2)$.

(2) $f_C(A, \textit{if } B) = f_C(B) \rightarrow f_C(A)$.

The notion of the satisfaction of an imperative is defined in terms of the truth of the propositional content of the imperative in a model M of the actual world:

(3) SATISFACTION
Imperative **I**, uttered in context C to addressee A, has been *satisfied* in a world represented by model M, if and only if A makes it the case that $f_C(\mathbf{I})$ is true in the world represented by M.

In words, an imperative has been satisfied if and only if its addressee made its content true. A first possible objection to this definition is that the mere fact that the addressee made the content true is to weak, especially, if we read 'satisfied' as 'complied with': consider a situation in which a person is told to get off a table and subsequently accidentally stumbles and falls off that table. In that situation, the propositional content of the imperative has been made true. Nevertheless, one would not want to claim that the person really complied with the imperative: his or her action was not influenced in the *appropriate way* by the utterance of the imperative clause.[2] The problem can be resolved by making a clear distinction between the notion of *satisfaction* of an imperative, and the notion of an addressee *complying with* an imperative. For the moment, let us assume that we are studying the former and look at another problem which casts a shadow over definition (3). Take a rule-like imperative such as:

(4) Make a cross, if you encounter a vampire.

In line with (2), the propositional content of this imperative is rendered by the following Predicate Logical sentence[3], assuming that the addressee's name is John:

(5) $encounter\_vampire(john) \rightarrow make\_cross(john)$

In words, if the addressee encounters a vampire, then he makes a cross. In Classical Logic the formulae of the form $p \rightarrow q$ and $\neg p \vee q$ (i.e., it is not the case that p or it is the case that q) express the same proposition: whenever one of them is true/false, the other one is as well. Thus, the propositional content of our imperative can also be expressed by the formula in (6).

(6) $\neg encounter\_vampire(john) \vee make\_cross(john)$

By combining this with the definition (3) we obtain: the imperative in (4) has been satisfied by John if and only if John makes least one of the following two formulae true:

(7) a. $\neg encounter\_vampire(john)$;
    b. $make\_cross(john)$.

According to this analysis, one possible way of satisfying the imperative would be for the addressee to simply avoid encountering a vampire (cf. Hamblin, 1987:84). In some sense, it is correct that in that case the imperative is satisfied (and even complied with), as predicted by definition (3). Still, I am left with the feeling (and hopefully the reader to) that the definition provides a very incomplete and potentially misleading analysis of the relation between imperatives and the actions which they give rise to.

Imagine that the addressee of the imperative is interviewed and asked why he was avoiding vampires. What could the addressee reply to this question? Would it, for instance, be appropriate for him to say: 'I avoided vampires because I was

[2] Suppose that the addressee of the imperative was a person named 'John' and the issuer a person named 'Mary'. In that situation, the following discourse would be, if not an infelicitous, certainly a misleading description of what took place: 'Mary told John to get of the table and he complied'. Note that in the aforementioned situation 'Mary told John to get of the table. He did not comply.' seems equally infelicitous/misleading, which suggests that 'A complied with imperative **I**' presupposes (rather than that it asserts) that A's action was influenced in the *appropriate way* by the utterance of **I**.

[3] For the moment, I am omitting an explicit representation of temporal reference.

told that if I see one, I should make a cross'? To me, this answer sounds odd if it is uttered out of the blue. It is however, not difficult to imagine a situation in which this answer makes sense. A hearer who assumes that the addressee for some reason or other wants to avoid making a cross will have no difficulty interpreting the answer: if the addressee of (4) was already committed to not making a cross, then he would indeed need to avoid vampires in order to comply with the imperative.

Generally speaking, one may conclude that in order to understand how imperatives influence the behaviour of their addressees the background information which also influences the actions of an addressee needs to be taken into account. In particular, one needs to take into consideration further propositions to which the addressee is committed. This means that although definition (3) provides us with some idea of the meaning of imperatives (in terms of their conditions of satisfaction), it fails to give us a full picture of how imperatives are *used* and (are expected to) *affect* the behaviour of their addressees.

3 Imperatives in Context: An Agent-based Analysis

In order to examine the influence of background information on the relation between imperatives and the actions which they give rise to, I am going to describe a class of models for communicating agents and their environment. The influence of background information can then be formalised in terms of constraints on the behaviour of the agents in such a model. These constraints will be normative in nature: they describe how the agents are *expected* to behave when they have agreed to comply with an imperative. Additionally, I will describe a concrete policy for action which allows an agent to fulfil these expectations.

At the basis of a Communicating Agents Model –a formally more explicit, but in some respects different account from the one advocated here[4], can be found in Piwek (2000)– lies a time line. This line consists of discrete instants of time and is infinite in both directions. In addition to the domain of instants of time (D_t) which underlies the time line, the model includes domains containing events (D_e), states (D_s) and individuals (D_i). The set of agents is a subdomain of the domain of individuals. The set of actions constitutes a subdomain of the domain of events. An event is an action if and only if it involves on or more agents who play the actor/agent role of the event (cf. Parsons, 1990).

Each of the members of the domains of events and states are event/state *tokens*. They can, respectively, *occur between* or *hold at* instants of time. I assume that if an event token occurs between two instants t_x and t_y (where x is precedes y), then it also occurs between all the instants t_a and t_b (where a precedes b) such that t_x precedes t_a and t_b precedes t_y. For instance, if somebody is cooking a meal between three and six 'o clock, then s/he is also cooking a meal between four and five 'o clock. Furthermore, I assume that for each event token there is a unique *maximal* pair of instants of time $\langle t_x, t_y \rangle$, i.e., interval, between which it

[4] In particular, the interpretation of the predicate *occur* is different.

occurs: any other pair between which the event occurs is covered by this maximal interval.

In this paper, I focus on a particular subclass of the subdomain of actions: the so-called *basic actions*. In AI planning theories these are usually conceived of as the leaves of hierarchical plans. They are the actions which an agent is supposed to directly execute: they do not permit decomposition into further constituent actions. In terms of the logical model which is proposed here, a basic action is an action whose occurrence cannot be entailed by the occurrence of another set of actions. Composite actions, on the other hand, are those actions whose occurrence is entailed by other actions. They do not exist independently of the actions by which they are entailed. An agent executes a composite action only indirectly by executing a number of basic actions which entail the occurrence of the composite action. Despite this elusive status of composite actions, agents can have propositional attitudes towards them: e.g., they can intend composite actions and devise plans to execute such actions. Usually such a plan consists of a sequence of (preferably basic) actions which entail the composite action. In the model-theoretic approach, which is adopted in this paper, the relation between basic and composite actions has to be formalized through axioms. These axioms stipulate which composite actions are entailed by combinations of basic actions.

We have seen that events occur between instants of time. We have also seen that for any event there is a maximal interval at which it occurs. In the context of this paper, it is helpful to have two-place relation $occur(e, t_x)$ which is interpreted as: event e starts to occur at the instant of time t_x. In other words, t_x is the beginning point of the maximal interval at which e takes place.

A state always holds *at* one or more instants of time. We write $hold(s, t)$ to say that state s holds at the instant of time t. For example, consider a situation in which an agent switches off the light in a room between the instants of time t_0 and t_1. This means that an event token e of the type *switch* took place at the transition from t_0 to t_1. Similar to Parsons (1990), I assume that an event can involve a number of objects: in this case, its actor (the agent who switches off the light) and its patient (the light which is switched off). Furthermore, there are two states s_0 and s_1: one in which the light is on and one in which the light is off. s_0 holds at $\ldots, t_{-2}, t_{-1}$ and t_0 whereas s_1 holds at $t_1, t_2, \ldots$ In general states are assumed to span a connected series of instants of time. In other words, if a particular state holds at two different instants of time t_x and t_y, then it also holds at all the instants between t_x and t_y. Additionally, a state (token) has the same type and involves the same objects at all the instants at which it occurs).

Agents are modelled as a subdomain of the domain of individuals. They are distinguished from other objects by their ability to carry (propositional) information and use that information to guide their actions. The role of information in the (dialogue) behaviour of agents has been studied extensively. See, for instance, Beun, (1992), Bunt & Katwijk (1979), the collection of papers in Bunt and Black (2000), Bunt et al. (this volume), Lewis (1979) and Power (1979). A number of notions, such as desire, belief, intention, plan, goal and commitment, have become central to such studies (see, also, Cohen et al., 1990; Grosz

& Kraus, 1996 and Lochbaum, 1994).

In this paper, I assume that imperatives give rise to commitments and focus on how the commitments of an agent determine which actions s/he carries out. In order to clarify the notion of commitment, it is useful to relate it to (my interpretation of) some of the other aforementioned notions.

Let us start with the notion of desire. Desires lie at the basis of an agent's behaviour. They are not necessarily rational: an agent can desire something which s/he believes to be impossible or have mutually contradictory desires. Goals are situations which an agent (1) dubs desirable, (2) believes to be possible, i.e., the situation is not inconsistent with his or her beliefs *and* other goals, and last but not least (3) has decided to realize. If a particular situation S_g is a goal for an agent, we also say that s/he intends that S_g. The relation between a desire and a goal can be fairly indirect, e.g., an agent A might adopt a particular goal because another agent proposed that goal and A has a desire to *cooperate* with that agent.

In order for an agent to realize his or her goals, s/he has to translate these goals into a set of basic actions which s/he can carry out and which lead to the goal situation. Such a set (usually a partially or totally ordered one) of basic actions is known as a *plan*. From a logical point of view, finding a plan can be seen as finding a description of a set of basic actions such that the description together with the agent's beliefs entails his or her goals:

(8) PLAN
Given an agent A, a set of beliefs B and a set of goals G of A, a plan P is description of a set of basic actions by A such that $B \cup P \models G$.

Here I want to suggest a natural generalization of this notion of a plan which my formulation of the notion of a plan in logical terms invites. Let us relax the constraint that a plan is a description of set of basic actions. We still assume that a plan needs to entail the goals of the agent. The plan itself is, however, defined as a description of the agent's (controlable) behaviour: it can describe basic actions by the agent, but also the disposition to abstain from certain actions or to carry out or abstain from certain actions in particular circumstances.

I propose to provide a formal definition of the notion of a generalized plan through a negative characterization. Basically, the idea is that a generalized plan should not entail any information to the effect that some state or event –except for a basic action– holds at/occurs between some instants of time or does not, *or* holds at/occurs between some instants of time or does not provided that certain conditions hold.[5]

(9) GENERALIZED PLAN
Given an agent A, a set of beliefs B and a set of goals G of A, a generalized

[5] The notion of a generalized plan is different from the notion of a *partial plan*, which is common in AI planning theories. A partial plan is a traditional hierarchical plan which has at least one leaf which is not a basic action. Partial plans, as opposed to generalized plans, do do not contain logically complex propositions.

plan is a description P such that $B \cup P \models G$ and there is no formula F of the form $p \rightarrow q$ or of the form q such that (1) $P \models F$, (2) F is not a tautology and (3) q describes that an event or state which is not a basic action by A holds/occurs or does not at or between some instants of time.

Let me explain why this generalization is useful by means of an example. Suppose an agent has the goal not to be bitten by a vampire. Furthermore, suppose that the agent believes that he will be bitten by a vampire if and only if he encounters a vampire and the vampire is not deterred. Finally, our agent knows that a vampire can be deterred by making a cross. A perfectly valid plan for this agent would be the plan to make a cross if he encounters a vampire. Such a plan is, however, not captured by the traditional notion of a plan as paraphrased in (8). The proposition *that the agent makes a cross if he encounters a vampire* does not describe a basic action. Rather, it describes a disposition.

Dispositions are covered by definition (9). The disposition *that the agent makes a cross if he encounters a vampire* together with the agent's beliefs entails that the agent will not be bitten, i.e., the first condition of (9). The definition contains further conditions to prevent agents from adopting plans which they cannot execute. For instance, an agent is not allowed to simply include in his or her plan that some event or state (which is not a basic action) will or will not occur/hold. Note that the condition also excludes as part of a plan formulae which say that: If state s holds then event e will occur, where e is not a basic action of the agent.

The notion of commitment and more specifically that of a commitment state of an agent still needs to be defined. In this paper the term commitment state is used to refer to the union of an agent's plans and his or her beliefs. This is the information which s/he is committed to.[6] The commitment state of an agent puts constraints on the behaviour of the agent. When we look at the model from the outside, we can speak of the satisfaction conditions of the commitment state. I assume that as an external observer we have access to a function *Int* which maps commitment states (of individual agents) onto Predicate Logical sentences. The conditions of satisfaction of a commitment state correspond with the truth-conditions of the corresponding Predicate Logical sentences in the model (which includes the agent and its commitment state).[7]

The reader might feel somewhat uncomfortable with the assumption that an observer has a function which returns the conditions of satisfactions of an agent's commitment state. It seems that we can only guess at such a function based on the behaviour of the agent, assuming that an agent strives for the satisfaction

[6] This definition of commitment deviates from the one given in Hamblin (1971). Hamblin is concerned with situations where an agent becomes publically committed to some piece of information. It is beyond the scope of this paper to delve into the relation between the private commitments which are discussed here, and Hamblin's public commitments.

[7] Technical details with respect to *Int* can be found in Piwek (2000). There, a partialized truth-conditional semantics is employed to avoid paradoxes such as 'The Liar'.

of his or her commitments. This is certainly correct when we consider human agents. However, the nice thing about building models is that we can simply define such a function for the agents with which we populate the model and then compare different policies for action with respect to this function. Thus we can determine how successful different policies are with respect to satisfying the agents' commitments. It will turn out that this perspective is very helpful for understanding the role of imperatives. Now that a model is in place, let us examine how we can use it to analyse the use of an imperative. Take the imperative (4), here repeated as (10).

(10) Make a cross, if you encounter a vampire.

The utterance event u of this imperative is modelled as occurring between two instants of time, say t_0 and t_1.[8] This utterance event has certain physical properties and some of these properties will be perceived and processed by the addressee of the utterance. In other words, in parallel with the utterance event, i.e., also between t_0 and t_1, there will be an observation event (by the addressee) of the utterance event. This observation event causes a change to the commitment state of the addressee. It is beyond the scope of this paper to provide a detailed analysis of how the perception affects the information state of the addressee (but see, for instance, Beun, 1994).

Let us assume that the addressee perceives and understands the imperative and believes that the utterance was felicitous (e.g., the speaker was sincere. See Austin, 1962). The commitment state of the addressee at t_1 will then contain the information of the addressee's commitment state at t_0 extended with the belief that it is the goal of the speaker that the addressee makes a cross if s/he encounters a vampire.[9] It is up to the addressee to signal whether s/he is committed to satisfying the speaker's goal. In other words, if the addressee signals compliance with the imperative, then s/he signals to have adopted the goal and also to have a plan for realizing it. The latter becomes a commitment of the addressee.

Thus, I assume that signalling compliance is stronger than merely adopting a goal, it also encompasses being committed to having a plan for that goal. Consideration of certain dialogue fragments seems to provide some support for this hypothesis. Consider:

[8] For the sake of the argument I assume that an utterance event consists of just one single transition between adjacent instants of time. In a more sophisticated analysis we would model an utterance as consisting of a composite event.

[9] In case the imperative is intended as advice, the speaker will have expressed the proposition that it is beneficial to the addressee to adhere to the propositional content of the imperative. This in contrast with a command where the speaker expresses his own goal or that of somebody different from both speaker and addressee on whose behalf the speaker issues the imperative. In this paper, such distinctions are not at issue. Rather, the main concern is the commitment which the addressee takes on when s/he decides to satisfy an imperative.

(11) A: Finish the paper before tomorrow.
B: Ok, but I might fail to do it.

Here, B's utterance appears to be incoherent. B cannot first signal compliance with the imperative and then suggest that s/he might fail to comply. Consider substituting 'I intend to' for 'ok' and then reconsider the dialogue fragment:

(12) A: Finish the paper before tomorrow.
B: (Well) I intend to, but I might fail.

There seems to be nothing wrong with B's reply. This suggests that the meaning of 'ok' cannot be identical to 'I intend to' and that a better paraphrase migth be 'I will'. In other words, the agent believes that s/he will satisfy the goal which is expressed by the imperative in virtue of the fact that s/he is committed to a plan which entails (given the agent's beliefs) the satisfaction of this goal.

Let us return to (10). The propositional content of this imperative is added to the commitment state of the addressee. For the sake of the argument I assume that making a cross is a basic action. This allows us to assume that the plan for satisfying the goal introduced by the imperative is identical to the propositional content of the imperative.[10] Formally, we have that $Int(c_{t_1}) = \Phi$ and $Int(c_{t_2}) = \Phi \wedge f_C(10)$.[11,12] I assume that the propositional content $f_C(10)$ corresponds with the following formula, where $occur(e,t)$ means that event e started at t:

(13) $(\forall v\, t\, t'\, e.(vampire(v) \wedge encounter(john, v, e) \wedge occur(e,t) \wedge successor(t', t) \wedge after(t, t_{utt})) \rightarrow (\exists e'.make\_cross(john, e') \wedge occur(e', t')))$

In words, if the addressee –John– encounters a vampire at a time t after the utterance time of the imperative t_{utt}, then he makes a cross starting at the instant of time which immediately succeeds t. Let us look at two possible constraints governing the role of imperatives. Firstly, one could consider an individual imperative and demand that its content is true in the world.

(14) CONSTRAINT: TRUTH-CONDITIONAL IMPERATIVE SATISFACTION
If $f_C(\mathbf{I})$ is the content of an imperative **I** uttered at some point in time, then $f_C(\mathbf{I})$ is true in the world.

We omit the condition that this content should be made true by its addressee. We address that issue in a moment. Let us first look at a related constraint which concerns the entire commitment state of an agent. We have seen that the

[10] For imperatives which involve composite actions, the plan or commitments which they give rise to cannot be identical to the propositional content of the imperative. The plan will have to entail the propositional content of imperative and satisfy the other conditions of (9).

[11] Where c_{t_x} stands for the commitments of the addressee at t_x.

[12] This representation does not allow for anaphoric reference to indefinite antecedents. For that we would need a dynamic system such as Discourse Representation Theory, see, Kamp & Reyle (1993).

content of an imperative is added to the commitment state of the addressee of an imperative. Thus we can define an alternative notion of satisfaction which pertains to such commitment states:

(15) Constraint: truth-conditional commitment state satisfaction
If c is the commitment state of an agent at some point in time, then $Int(c)$ is true in the world.

The addressee of an imperative is supposed to actively make sure that the constraint (14) is satisfied. An agent who manages to live up to this expectation is said to have *complied with* the imperative. However, an agent who tries to comply with an imperative cannot address that imperative in isolation from his or her other commitments. Consider, for instance, our agent John. Suppose he is committed to encountering a vampire, because he replied 'ok' to the imperative 'Find a vampire'. This commitment prevents John from satisfying (10) by avoiding vampires. The question then is how an agent can make sure that all his or her commitments are true in line with (15), if it is not possible to consider them one at a time. *Is there a policy which allows an agent to make sure that all his or her commitments are/become true?*

Let us first try to formulate a more precise version of this question. At the basis of this section lies the idea of a model of reality. The models which we have described up till now are total in nature: such a model determines completely which states and events take place at any given instant of time/interval and who and what are involved. With respect to such a model we can check whether constraints such as (14) and (15) are satisfied. Consider a model which is just as determined as the aforementioned models except for the fact that the actions of a particular agent, say John, are not specified in this model at any interval. A policy provides us with a rule which takes such a partial model and turns it into a fully determined, i.e., total, one. Thus the policy determines which actions an agent does and does not carry out at any given instant of time.

Whether a policy is successful can now be decided by checking whether the total model which it delivers always satisfies the constraint (15). In other words, a successful policy is one which allows an agent to make sure that all his or her commitments are true.

(16) Policy for Action: First Version
If at time t agent A's commitment state is c and $Int(c)$ entails that A carries out some action a at t, then A should carry out a at t. Furthermore, A should only carry out actions which are required by this policy.

For now I have assumed that the normative constraint (15) is the only constraint on our models. At the end of this paper some further constraints are discussed, for instance, physical constraints to which any real-world events and states are subjected. Such constraints affect what actions an agent can perform at a given instant of time. A succesful policy should, of course, only suggest actions which are compatible with this kind of hard constraints.

Let us put this complication aside and look at some of the concrete predictions which the proposed policy entails. Consider an agent whose commitment state consists only of the following commitment:

(17) $(\forall v\, t\, t'\, e.(vampire(v) \wedge encounter(john, v, e) \wedge occur(e, t) \wedge successor(t', t) \wedge after(t, t_{utt})) \rightarrow (\exists e'.make\_cross(john, e') \wedge occur(e', t')))$

According to the policy (16), this agent will undertake no actions at all. This explains why we consider it strange for an agent to avoid vampires solely on the basis of the imperative (10). Alternatively, suppose that John's commitment state corresponds with the following formula:

(18) $vampire(v_1) \wedge encounter(john, v_1, e_1) \wedge occur(e_1, t_3) \wedge (\forall v\, t\, t'\, e.(vampire(v) \wedge encounter(john, v, e) \wedge occur(e, t) \wedge successor(t', t) \wedge after(t, t_{utt})) \rightarrow (\exists e'.make\_cross(john, e') \wedge occur(e', t')))$

In words, John is committed to v_1 being a vampire and encountering him at t_3. John's commitment state entails that at t_4 John makes a cross. And hence, according to policy (16), John should make a cross at t_4.[13] Finally, consider a situation in which John is committed to not making a cross:

(19) $\neg\exists e\, t.(make\_cross(john, e) \wedge occur(e, t)) \wedge (\forall v\, t\, t'\, e.(vampire(v) \wedge encounter(john, v, e) \wedge occur(e, t) \wedge successor(t', t) \wedge after(t, t_{utt})) \rightarrow (\exists e'.make\_cross(john, e') \wedge occur(e', t')))$

This commitment state entails, as is to be expected, that John should avoid vampires:

(20) $\forall v\, t\, e.(vampire(v) \rightarrow \neg(encounter(john, v, e) \wedge occur(e, t) \wedge after(t, t_{utt})))$

Let us look more closely at how successful the policy (16) is in making sure that an agent satisfies all his or her commitments. It turns out that we need to make a number of further assumptions to guarantee its success.

First, we need to impose a constraint on how the commitment state of an agent changes over time. It should not be possible that an agent has a commitment state c_{t_x} at t_x and c_{t_y} at t_y such that c_{t_x} and c_{t_y} are not consistent with each other. It is impossible to satisfy both states in one and the same world/model and therefore (15) would be violated.[14]

Second, if a commitment state c_{t_y} entails an action by an agent at t_x, then c_{t_x} should also entail this action. Thus, in line with the policy (16), the agent will actually carry out the action at t_x. Without this assumption, s/he would

[13] In fact, the commitment state also needs to include the information that t_4 is the successor of t_3. We assume that the commitment state alway contains necessary information to derive that one instant succeeds another.

[14] Note that the proposed assumption automatically rules out that an individual commitment state is inconsistent.

not do so, since according to the policy an agent only carries out the actions at time t to which s/he is committed at time t.

Third, we need to assume that we are only considering the actions which the agent itself is supposed to carry out. Of course, s/he might also have commitments pertaining to states, events and actions by others. However, for the moment, we need to assume that these commitments are by definition true in the model.

Fourth, we need to assume that the agent's commitment state at each point in time is *complete*. For instance, suppose an agent is only committed to (17). Furthermore, suppose that in that world the agent encounters a vampire at t_4, but does not realize this and therefore is not committed to encountering the vampire. In that case, according to policy (16) the agent will not make a cross. But, of course, he should have done so in order to satisfy constraint (15). In other words, for the policy to be successful the agent's commitment state should be complete in the following sense: there should be no true information which is not part of an agent's commitment state, but which if it were, would cause the agent to perform an action which s/he would not undertake if this information was not present.

Finally, we need to assume that the actions of an agent are independent of each other or that if there are any constraints governing the interactions between different actions, then these should be included in the commitment state of the agent. For instance, suppose that there is a constraint which says that if an agent does not do action a_1 at t_1, then he cannot do a_2 at t_2, such as: one cannot submit a paper without first writing it. Suppose an agent is committed to submitting a paper at t_2. If this were the only commitment that the agent maintained, then policy (16) would tell him/her to do nothing at t_1. But the aforementioned constraint would then prevent him/her from submitting a paper at t_2. Thus, the agent would not be able to live up to its commitments. To avoid such situations, an agent would need to be committed to the information expressed by the constraint. Suppose the agent were committed to the information that (1) if you do not write a paper at t_x, then you do not submit a paper at t_{x+1} and (2) the agent submits a paper at t_2. These two propositions together entail that the agent writes a paper at t_1. And, hence, at t_1 the agent has to write a paper according to (16).

According to (16) an agent will not carry out any actions which are not licensed by (16). This puts a serious constraint on the behaviour of an agent. The condition is intended to prevent an agent from performing some action which prevents him or her from carrying out some other action which s/he is committed to at a later time. For instance, one might, just for the fun of it, throw away a key, which one needs later on to open a lock. However, if we assume that an agent's commitment state captures such regularities (i.e., that the lock cannot be opened without the key), then we can formulate a more liberal policy:

(21) Policy for Action: Second Version

If at time t agent A's commitment state is c and $Int(c)$ entails that A carries out some action a at t, then A should carry out a at t. Furthermore,

A should only carry out actions which if they were to be added to his or her commitment state would not lead to an inconsistent commitment state.

I have presented a policy and a modified version of it. Both are only successful with respect to (15) when a number of further conditions/constraints hold. Many of these constraints may be seen as too restrictive for a real-world agent. However, I hope that by making these constraints explicit we have gained a starting point for the systematic study of agent behaviour. The next step would be to investigate how an agent can make sure that as much of the constraints as possible are satisfied, for instance, by means of an active policy for gathering information, such that all true information which is relevant to the commitments is available to the agent (in other words, a policy for maintaining a complete commitment state).

I conclude this section by drawing attention to a further problem in the analysis of imperatives which goes back to Ross (1941) and which has become known as Ross' Paradox. Ross noted that whereas p entails $p \vee q$ in classical logical, the imperative $p!$ does not entail the imperative $p \vee q!$ For instance, when I tell you to *Post a letter*, then I do not tell you to *Burn it or Post it.* In my approach, we maintain the Classical Logical inference from *Somebody posts a letter* to *Somebody posts a letter or burns it.* However, how information affects the behaviour of an agent is regulated by (21). An agent who is committed to *Posting a letter*, as a result of accepting the imperative 'Post the letter', will in line with (21) simply post the letter, since this is a (basic) action which is entailed by his or her commitment. The entailed proposition that s/he post or burn it has no effects on his or her actions. Note that if it did, this could lead to a violation of (15). Suppose the agent had a policy of also making complex sentences true directly. Then, s/he might choose to make the disjunction true by burning the letter. But this would prevent him or her from satisfying the commitment to posting the letter.

Alternatively, a person who accepts to *Burn or post a letter* will have a commitment state which contains this proposition. From this proposition the agent can infer no action whatsoever. Here we need to invoke the completeness assumption to get the agent moving. This assumption says that there should be no true information which is not part of an agent's commitment state, but which if it were, would cause the agent to execute an action which s/he would not undertake if this information was not present. Suppose our agent considered doing nothing at first. That would mean that s/he would not burn the letter. But this information together with the commitment to burn or post it entails that the agent should post it. Alternatively, if the agent were to first consider not posting the letter, this would cause him or her to burn it. If we invoke the completeness assumption, then it turns out that that the agent either has to post the letter or burn it.

4 Conclusions and Further Issues

An analysis of imperatives has been provided which takes into account the influence of background information on the use of (complex) imperatives. In particular, I have described a model of communicating agents in which imperatives modify the commitments of these agents. I proposed to understand an agent's commitments as his or her beliefs combined with his or her (generalized) plan for action. The notion of a generalized plan was introduced as a natural extension of the traditional notion of a plan when it is looked at from a logical perspective.

The commitments are expected to influence the behaviour of the agent. I formulated a normative constraint which required that the commitments of the agent should be true. Subsequently, I examined a policy which an agent might adopt in order to satisfy this constraint. It turned out that various further practical constraints are needed to make the policy successful, i.e., to guarantee that all of the agents commitments are made true.

The proposed analysis of imperatives uses Classical Predicate Logic to represent the propositional content of imperatives. I showed that such an analysis does not need to run into certain problems which have been associated with Classical Logical analyses of imperatives. Therefore, we did not depart from Classical Logic.[15] In this respect, my analysis of imperatives is Gricean in nature (Grice, 1975): we retain the Classical Logical core of the analysis and account for certain features of the use of imperatives in terms of rules which govern the relation between the content/information which is conveyed by imperatives and the role which this information plays in the behaviour of the addressee. In other words, we make a distinction between the questions 'What does the propositional content of an imperative entail?' and 'What can we conclude when somebody utters an imperative with a particular content?'.

The work reported in this paper can be seen as complementary to much work in AI on planning which focuses on commitments to actions which are complex in the sense that these actions can be decomposed into a collection of sub-actions. The actions in question are, however, typically not part of a plan which is logically complex. In this paper, *logically* complex commitments, which include the plans of an agent, have been at the centre of our investigation. Due to the focus in AI on logically simple plans, much work in that tradition pays little attention to the relation between an agent's plan and the actions which s/he carries out on the basis of that plan. However, when we move to logically complex plans/commitments, the relation between an agent's actions and his plans/commitments is no longer trivial, and it is exactly that relation which has been studied in this paper.

Finally, let me indicate some issues for further investigation. There is, for instance, the inherent temporal vagueness of imperatives which has not been addressed in this paper. Consider the simple imperative 'open the door'. A paraphrase of this imperative might be: *Open the door as soon as possible after now but not too long after now.* It is not immediately clear how this paraphrase can

[15] As is, for instance, advocated in Perez Ramirez (2000).

be expressed by means of a Predicate Logical formula. Furthermore, a shortcoming of the current model is that it does not cater for an agent who happens to realize that it violated one of its commitments. Suppose the commitment was p. The realization that p is not satisfied should result in the addition of $\neg p$ to his or her commitment state. But clearly, a strictly additive update with this information leads to an inconsistent state. In short, a full-fledged theory of imperatives presupposes a theory of commitment revision.

Acknowledgments

The problems concerning the relation between imperatives and action first attracted my attention when working, from 1994 until 1998, on the DenK project (see Bunt et al., this volume). I would like to thank my colleagues on the DenK project for the stimulating environment which put me on a track which eventually led to the work which is reported in this paper. Furthermore, thanks are due to Robbert-Jan Beun, Kees van Deemter and Miguel Perez Ramirez for commenting on earlier versions of this paper. Naturally, I take responsibility for any remaining mistakes in this paper.

References

Austin, J. (1962) *How To Do Things With Words.* Oxford: Clarendon Press.

Beun, R.J. (1992) A framework for Cooperative Dialogue. In: Taylor, M., F. Néel & D. Bouwhuis (eds.), *The Structure of Multimodal Dialogue II*, Maratea, Italy: Proceedings of the Venaco Workshop on Multimodal Dialogue.

Beun, R.J. (1994) Mental state recognition and communicative effects. *Journal of Pragmatics* **21**, 191–214.

Biber, D., S. Johansson, G. Leech, S. Conrad & E. Finegan (1999) *Longman Grammar of Spoken and Written English.* Harlow: Longman.

Bunt, H. & A. Van Katwijk (1979) Dialogue acts as elements of a language game. In: W. Zonneveld & F. Weerman (eds), *Linguistics in the Netherlands 1977–1979*, Dordrecht: Foris Publications, 264–282.

Bunt, H. & W. Black (eds.) (2000) *Computational Pragmatics: Abduction, Belief and Context.* Amsterdam/Philadelphia: John Benjamins.

Cohen, P., J. Morgan & M. Pollack (1990) *Intentions in Communication.* Cambridge Massachusetts: The MIT Press.

Grice, H. (1975) Logic and Conversation. In: P. Cole & J. Morgan (eds.), *Speech Acts. Syntax and semantics*, Vol. 3. 41–58, New York: Academic Press.

Grosz, B. & S. Kraus (1996) Collaborative Planning for Complex Group Action. *Artificial Intelligence* **86**(2), 269-357.

Hamblin, C. (1971) Mathematical Models of Dialogue. *Theoria* **37**, 130–155.

Hamblin, C. (1987) *Imperatives.* Oxford: Blackwell.

Kamp, H. & U. Reyle (1993) *From Discourse to Logic.* Dordrecht: Kluwer Academic Publishers.

Lewis, D. (1979) Scorekeeping in a Language Game. In: Bäuerle, R., U. Egli & A. von Stechow (eds.), *Semantics from Different Points of View*, Berlin: Springer 172–187

Lochbaum, K. (1994) *Using Collaborative Plans to Model the Intentional Structure of Discourse*. PhD. thesis, Cambridge, Massachusetts: Harvard University

Parsons, T. (1990) *Events in the Semantics of English.* Cambridge, Massachusettes: The MIT Press.

Perez Ramirez, M. (2000) Imperatives, State of the Art. In: Kilgarriff, A. et al. (eds.), *Proceedings of CLUK 3*, Universities of Brighton and Sussex, 18–25.

Piwek, P. (2000) Imperatives, Commitment and Action: Towards a Constraint-based Model. In: *LDV Forum*, Special Issue on Communicating Agents (To appear). Available at: `http://www.itri.bton.ac.uk/~Paul.Piwek`

Power, R. (1979) The Organisation of Purposeful Dialogues. *Linguistics* **17**, 107–152.

Ross, A. (1941) Imperatives and Logic. *Theoria* **7**, 53–71.

Searle, J. (1969) *Speech Acts: An Essay in the Philosophy of Language.* Cambridge: Cambridge University Press.

Stenius, E. (1967) Mood and language-game. *Synthese* **17**, 254-274.